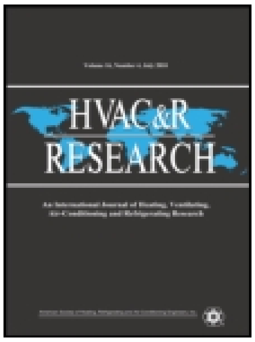


HVAC&R Research

ISSN: 1078-9669 (Print) 1938-5587 (Online) Journal homepage: http://www.tandfonline.com/loi/uhvc20

Calibrating Detailed Building Energy Simulation Programs with Measured Data—Part II: Application to Three Case Study Office Buildings (RP-1051)

T. Agami Reddy , Itzhak Maor & Chanin Panjapornpon

Calibrating Detailed Building Energy Simulation Programs with Measured Data—Part II: Application to Three Case Study Office Buildings (RP-1051)

T. Agami Reddy, PhD, PE Member ASHRAE
Itzhak Maor, PhD, PE Member ASHRAE
Chanin Panjapornpon, PhD

Received December 21, 2005; accepted July 31, 2006

The companion paper proposed a general methodology of calibrating detailed building energy simulation programs to performance data that also allowed the determination of the prediction uncertainty of intended energy conservation measures. The methodology strived to provide a measure of scientific rigor to the process of calibration as a whole, which has remained an art form with no clear consensus guidelines despite being followed by numerous professionals for several decades. The proposed methodology, while providing a clear structure consistent with that adopted in more mature scientific fields, also uses expert domain knowledge and is flexible enough to satisfy different users with different personal preferences and biases. This paper attests to the overall validity of the methodology by presenting the results of applying it to three case study office buildings—two synthetic and one actual. Conclusions on various variants of the overall calibration methodology are presented, along with guidelines and a summary of lessons learned on how to implement such a calibration methodology. Future research needed prior to implementation in commercial hourly detailed simulation programs is also identified.

INTRODUCTION

A previous paper (Reddy 2006) provided a pertinent and detailed literature review of calibrated simulation techniques, describing their uses, strengths, weaknesses, procedures, and tools as well as pertinent issues related to model fitting uncertainty. The companion paper (Reddy et al. 2007) proposed a general methodology of calibrating detailed building energy simulation programs to performance data that consisted of various concepts adapted from the general scientific literature. The calibration methodology involved the following five major steps:

1. An important first step is to prepare a preliminary simulation input file of the building that is as realistic and error-free as possible. This would entail making sure that the simulation program has the capability of modeling the type of building and systems present and that the inputs have been entered correctly.
2. Next, reduce the dimensionality of the parameter space by performing walk-through audits and heuristics. For a given building type, identify/define a set of influential parameters and building operating schedules along with their best-guess estimates (or preferred values) and their range of variation characterized by either the minimum-maximum range or the upper and lower 95th probability threshold values. The set of influential parameters to be

T. Agami Reddy is a professor in the Department of Civil, Architectural and Environmental Engineering, Drexel University, Philadelphia, PA. **Itzhak Maor** is a principal at PWI Energy Inc., Philadelphia. **Chanin Panjapornpon** is an assistant professor in the Department of Chemical Engineering, Kasetsart University, Bangkok, Thailand.

selected should be such that they correspond to specific and easy-to-identify inputs to the simulation program.

3. Next, perform a "bounded" coarse grid calibration (or unstructured or blind search) using a mid-point Latin Hypercube Monte Carlo (LHMC) simulation involving numerous trials or realizations with different combinations of input parameter values. This would allow filtering or identification of the most promising solutions of input parameter vectors and also provide a means of identifying the more sensitive or strong parameters by performing a regional sensitivity analysis (RSA).
4. Subsequently, perform a guided search calibration to further refine or improve the calibrated solutions identified by the coarse grid search.
5. Rather than using only the best calibrated solution (determined on how well it fits the data) to make predictions about the effect of intended energy conservation measures (ECMs), use a small number of the top plausible solutions. Not only is one likely to obtain a more robust prediction of the energy and demand reductions, but this would allow determining their associated prediction uncertainty as well. The justification of doing this has been pointed out in Reddy et al. (2007) but is repeated below. The conventional wisdom was that once a simulation model was calibrated with actual utility bills, the effect of different intended ECMs could be predicted with some degree of confidence by making changes to one or more of the model input parameters that characterize the ECM. Such thinking is clearly erroneous since the utility billing data are the aggregate of several end-uses within the building, each of which is affected by one or more specific and interacting parameters. While performing calibration, the many degrees of freedom may produce good calibration overall even though the individual parameters may be incorrectly identified. Subsequently, altering one or more of these incorrectly identified parameters to mimic the intended ECM is very likely to yield biased predictions.

Though the methodology is applicable to any building energy simulation program, the scope of ASHRAE Research Project 1051 was restricted to the DOE-2 program, a widely used, public domain, fixed schematic hourly simulation program (Winkelmann et al. 1993) and to the widely prevalent case where year-long utility billing data are the only performance data available for calibration. It was presumed that the level and accuracy of knowledge about the building geometry, scheduling, and various system equipment would be consistent with a "detailed investment grade" audit, involving equipment nameplate information as well as some limited on-site measurements (clamp-on meters, etc.) performed during different times of the day (morning, afternoon, night) as well as over different days of the week in order to better understand variability in some of the simulation inputs. The intent of this paper is to summarize the results of applying the calibration methodology to two synthetic and one actual office buildings, as well as draw practical guidelines and conclusions on how to practically implement such a calibration methodology. Detailed descriptions of the buildings and systems (as well as the architectural and system files input to the simulation) can be found in Reddy et al. (2006).

DESCRIPTION OF BUILDINGS AND PARAMETERS

Two types of generic office buildings have been selected: (a) simpler buildings, generally small to medium in size with two or three stories, using unitary-type equipment with one source of energy, namely, electricity only, and (b) more complex buildings, generally medium to large in size with several stories, using decentralized equipment and more than one source of heat, namely, electricity and gas.

The calibration methodology was evaluated against two synthetic buildings (one simple and one complex) and a third actual building, which falls into the simple building category. Evaluation using the synthetic buildings involved selecting a building and specifying its various construction and equipment parameters as well as its operating schedules (called *reference values*), and using the DOE-2 simulation program to generate "electric utility bill" data for a whole year coinciding with calendar months. The utility billing data are then assumed to be the measured data against which calibration is performed. Since the "correct" or reference parameters are known beforehand, we can evaluate the accuracy and robustness of the proposed calibration methodology by determining how correctly the calibrated models can fit the utility bill data and also how accurately the effect of various ECMs can be predicted.

For the sake of simplicity, the calibration is assumed to be performed for the case when exactly one year of utility bills is available (i.e., 12 monthly bills). In reality, the analyst may use multi-year utility bills provided the building has not undergone any major changes in operation and equipment retrofits/replacements, since this is likely to reduce calibration uncertainty and result in a more robust calibration. This aspect is left for a follow-up study to investigate.

A summary of key information related to building size, geometry, and systems assumed for all three buildings is provided in Table 1. Synthetic building S1 is assumed to be a typical single-story office building of 25,000 ft^2 (2,322 m^2) with packaged variable air volume (VAV) and electric heat representative of a simple building with one source of energy (electricity only). Typical Meteorological Year (TMY2) data from two geographic locations were selected for evaluation, Philadelphia, PA, and Dallas, TX. The synthetic complex office building selected (S2) is also simulated for two weather conditions, Atlanta, GA, and Dallas, TX, using TMY2 data. It is a class A large building (218,000 ft^2 [20,289 m^2]) with seven floors and a penthouse with the lobby, cafeteria, service areas, and mechanical/electrical rooms on the first floor and offices on the remaining floors. Building cooling is provided by electricity, while heating is by natural gas. The actual office building selected (A1), located in Collegeville, PA (about 20 miles west of Philadelphia), is a class A building constructed in 2002 and occupied since the summer of 2003. Though a large building with four floors, it is all-electric and has four VAV rooftop units. Utility bills for a whole year (January–December 2004) are used as the basis for the calibration, along with included pertinent billing period read dates (which are around mid-calendar month), number of billing days, total electrical consumption, and demand data. The total electrical consumption was normalized to a daily basis, which is the values used during the calibration process. Weather data (temperature, solar, wind, etc.) synchronous with the utility billing periods are required by the simulation program and were picked from hourly climatic data acquired from the Climatic Services Division of the National Climatic Data Center, Ashville, NC.

Architectural details and summary sheets of the building, zones, systems, and plant produced by the DOE-2 simulation program, as well time series and scatter plots of the month-by-month variation of billing data (kWh, kW, and gas usage) for all three buildings, are fully described in Reddy et al. (2006). This report also details the type of primary and secondary systems, hot water system, lighting, controls, and occupancy schedules. For the actual building, much of these data were supplied by the building manager as well as from a site visit. Finally, for the sake of practicality, we have used floor-weighted weighting factors (rather than custom weighting factors) and also selected suitable default equipment performance curves (fan, chiller, boiler, etc.) that are available in DOE-2.

Note that the actual electric demand is usually charged based on either 15 minutes or 30 minutes usage. However, most of the building energy simulation programs implicitly assume hourly simulation time steps, so reconciliation of the measured and simulated demand contains this inherent mismatch, which introduces a certain amount of error during calibration.

Table 1. Description of Building and Systems Assumed for the Two Synthetic (S1 and S2) and One Actual (A1) Office Buildings

	Office	Building S1	Building A1	Building S2
General	Total size, ft^2 (m^2)	25,000 (2,322)	108,576 (10,087)	218,400 (20,289)
	No. of stories	Single	3	7 + Penthouse
	Dimensions (plan), ft (m)	250 × 100 (76.2 × 30.5)	261 × 104 (79.6 × 31.7)	330 × 90 (100.6 × 27.4)
	No. of zones	5	10	12
Building Envelope	Roof	4 in. (10 cm) concrete, 1 in. (2.5 cm) insulation	4 in. (10 cm) concrete, 3 in. (7.5 cm) insulation	4 in. (10 cm) concrete, 3 in. (7.5 cm) insulation
	Wall U-factor, Btu/h·ft^2·°F (W/ K·m^2)	0.100 (0.568)	0.07 (0.40)	0.30 (1.70)
	Window U-factor, Btu/h·ft^2·°F (W/K·m^2)	0.57 (3.24)	0.35 (1.99)	0.65 (3.69)
	Shading coefficient (SC)	0.75	0.48	0.8
Internal	Lighting power density, W/ft^2 (W/m^2)	1.7 (18.3)	1.5 (16.1)	1.8 (19.4)
	Equipment power density, W/ft^2 (W/m^2)	1.0 (10.8)	1.15 (12.4)	0.9 (9.7)
	Occupant density, ft^2/person (m^2/person)	200 (18.6)	406 (37.7)	275 (25.5)
Ventilation	Occupied/unoccupied, cfm/person (L/person)	15/0 (7.08/0)	15/0 (7.08/0)	15/0 (7.08/0)
Thermostat Setting	Occupied, °F (°C)	72 /21	72 /21	75 /21
Primary Equipment		Rooftop	Rooftop	WC, centrifugal chillers, gas powered HW boilers
	Size: Cooling, tons (kW) Heating, MBtu (MJ)	N/A	N/A	700 (2,450) 6.0 (6.3)
	Energy efficiency: cooling/heating	N/A	N/A	0.72 kW/ton 80%
Secondary Equipment	Type	PVAV	PVAV	VAV + HW reheat
	Size (tons)	Cooling-70	Cooling-310	N/A
	Energy efficiency	8.0 EER	9.0 EER	N/A
Heating	Electric/fossil	Electric	Electric	HW/natural gas

HEURISTIC INFLUENTIAL INPUT PARAMETERS

Table 2 assembles a list of heuristically identified influential parameters that have simple and clear correspondence to specific and easy-to-identify inputs to the DOE-2 simulation program. The simple building category includes 20 parameters, while the complex building has 4 parameters (P21–P24) relating to primary heating and cooling plants in addition to the same 20 parameters. Note that the list consists of seven discrete schedules, thirteen continuous parameters, and four binary parameters (i.e., either ON or OFF or one of only two possible categories). Table 2 also contains information about the range of the 24 parameters (i.e., the minimum and maximum values the parameter can assume) as well as the base or preferred values, i.e., the values of the various parameters that the analyst deems most likely. It should be pointed out that the discrete parameters P1–P7 are diurnal schedules that consist of a set of 24 diurnal values. The hour-by-hour numerical values of these seven parameters are given both as tabular data and as graphs in the report by Reddy et al. (2006). Note that three sets of values and graphs have been stipulated for each variable: minimum, base, and maximum for weekdays only, Saturdays, and Sundays/holidays. As an illustration, Figure 1 depicts the three assumed schedules for the heating schedule (parameter P5) during weekdays.

COARSE GRID SEARCH

As fully described in the companion paper (Reddy et al. 2007), the intent of the coarse grid search process is to *reduce the dimensionality* of the calibration problem by identifying strong

Table 2. List of Influential Parameters for the Complex Office Building[a]

No		Description	Variable Type	Unit	Minimum value	Base value	Maximum value	Reference Case value
1	Load Schedules (Rooms)	Lighting Schedule	D	NA	OffLgt_1	OffLgt_D	OffLgt_2	OffLgt_D
2		Equipment Schedule	D	NA	OffEqpt_1	OffEqpt_D	OffEqpt_2	OffEqpt_2
3		Auxiliary Equipment Schedule	D	NA	AuxOffEqpt_1	AuxEqpt_D	AuxOffEqpt_2	AuxEqpt_1
4	Systems Schedule (Zones)	Fans Schedule	D	NA	OffFan_1	OffFan_D	OffFan_2	OffFan_2
5		Space Heating Temperature Schedule	D	NA	OffHtT_1	OffHtT_D	OffHtT_2	OffHtT_1
6		Space Cooling Temperature Schedule	D	NA	OffClT_1	OffClT_D	OffClT_2	OffClT_D
7		Outside Air (Ventilation) Schedule	D	NA	OffOA_1	OffOA_D	OffOA_2	OffOA_2
8	Envelop Loads	Window Shading Coefficient	C	Fraction	0.16	0.75	0.93	0.8
9		Window U value	C	Btu-h/Deg F*Sqft	0.25	0.57	1.22	0.65
10		Wall U value	C	Btu-h/Deg F*Sqft	0.0550	0.1000	0.5800	0.3000
11	Internal Loads (Rooms)	Lighting Power Density	C	W/Sqft	1.3	1.7	2	1.8
12		Equipment Power Density	C	W/Sqft	0.8	1.0	1.2	0.9
13	Systems Variables	Supply Fan Power/Delta_T	C	Kw/CFM	0.00124	0.00145	0.00166	0.00135
14		Energy Efficiency Ratio (EIR)	C	Fraction	0.1564	0.1849	0.2275	0.20478
15		On Hours Control	D	NA	VFD	IGV	IGV	IGV
16		Off Hours control	D	NA	OFF	Cycle on any Zone	Cycle on any Zone	Cycle on any Zone
17		Minimum Supply Air Flow	C	Fraction	0.3	0.65	1.0	0.7
18		Economizer	D	NA	Yes	Yes	No	Yes
19		Minimum Outside Air	C	Fraction	0.1	0.3	0.5	0.4
20	Auxiliary Electrical Loads	Auxiliary Electrical Loads -Non- HVAC effect	C	KW	25	50	75	65
21		Cooling Tower Fan power	C	BHP/GPM	0.0118	0.0184	0.0212	0.0189
22		Cooling Tower Fan Control	D	NA	VFD	Two Speed	Single Speed	Two Speed
23		Primary CHW & Cond.Pump Flow	C	GPM/Ton	2.25	2.7	3.38	2.9
24		Boiler Efficiency Ratio	C	Fraction	1.25	1.43	1.54	1.33

a. The minimum, base (or best guess), and maximum values characterize the assumed range, while the reference values are those used to generate the synthetic utility bills used for calibration. The simple office building category does not include the last four parameters related to primary systems.

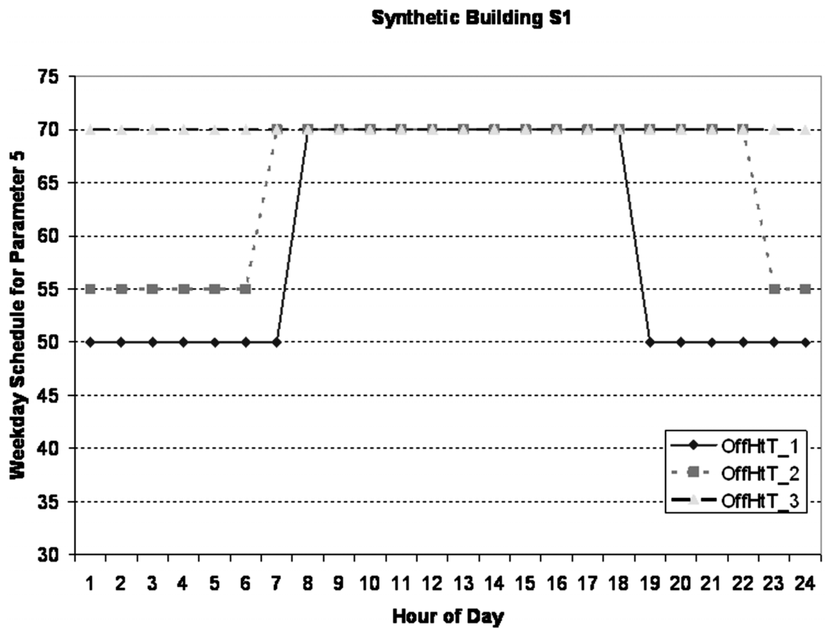


Figure 1. Three discrete space heating temperature schedules (parameter P5) during weekdays assumed during the coarse grid search. Similar discretization has been specified for weekends and holidays for this parameter as well as the other discrete parameters listed in Table 2.

and weak variables (the remainder being termed *uncertain*) among the set of influential parameters described above as well as to identify the sub-ranges most likely to contain the "actual" value of each of the strong parameters (recall that the range of variation of each parameter is discretized into three sub-ranges). This is to be achieved by adopting a blind LHMC search coupled with a RSA. In general, Monte Carlo (MC) filtering is the process of rejecting sets of model simulations that fail to meet some prescribed criteria of model performance (such as the goodness-of-fit [GOF] criteria suggested in the companion paper (Reddy et al. 2007) based on the normalized mean bias error (NMBE), the coefficient of variation (CV), or a combined index that weights them both (total). Sensitivity analysis goes a step further in that it allows identifying the parameters that are strong and those that are weak among the influential parameter sets that meet the prescribed satisfactory criteria. Once these are determined, the values of the weak parameters can be fixed at their nominal values, thus reducing the number of parameters to be calibrated. Once an LHMC batch run is completed, the GOF_CV and GOF_NMBE indices are computed for each trial, from which those parameter vectors that result in high GOF numbers (i.e., those whose predictions fit the utility bills poorly) can be weeded out. The information contained in these "good" or promising parameter vectors can be used to identify the weak parameters, which can then be removed from further consideration in our calibration process. This is a type of sensitivity test where the weak and strong parameters are identified using non-random pattern tests (Saltelli et al. 2000; Saltelli 2000). The Chi-square (χ^2) test was advocated as a test for statistical independence for each and every parameter.

It must be pointed out that LHMC is a relatively efficient sampling process. For example, we have 20 parameters (p) with each at three different levels, then the number of combinations

$M = 3^{20} \approx 3.5 \times 10^9$. Even reducing p by five parameters would result in $M = 3^{15} \approx 1.4 \times 10^7$, which is more than a two orders of magnitude reduction. Hence, reducing the dimensionality of the search is of critical importance. Note that with $p = 20$, performing even 3500 batch simulation trials would result in sampling the search space once in a million possibilities, which is a very sparse sampling process. In an attempt to make the LHMC process more efficient, we evaluated two types of tests: (a) single-stage tests, where a large number of LHMC tests are performed all together in one batch run, and (b) two-stage (or higher) tests, where a smaller number of tests are run and the results are analyzed to determine weak parameters whose values are frozen; then a second number of runs are run. Higher stages would involve repeating the iteration until no further weak parameters are identified. The multi-stage process may be more computationally efficient but would require analyst intervention at each stage.

Numerous calibration runs were performed for all three buildings (described in Reddy et al. 2006), while for the sake of brevity only a representative selection of results is assembled in Table 3. There are seven sets of results, three for S1 (two for the building assumed located in Philadelphia and one for Dallas), two for A1, and two for S2 (one for Atlanta and one for Dallas). Set 1 and Set 2 correspond to building S1 with base values of the 20 influential parameters *intentionally* chosen to be equal to the reference values. This idealized state would provide an upper limit to the calibration accuracy that one could hope to achieve. Such a simplification has not been done in all the remaining synthetic cases, i.e., sets 3, 6, and 7. Sets 4 and 5 correspond to A1, with the difference being that the ranges of some of the parameters have been readjusted, an issue that is described below. The LHMC runs have been coded "a" to denote one-stage calibration and "b" to denote two-stage calibration. The number of trials for each run as well as the minimum, maximum, and median of the top 20 calibrated runs in terms of GOF_Total are shown in the table. Equal weighting factors for kWh/kW/gas utility bills are assumed, while a ratio of 3:1 has been selected for NMBE:CV values. Calibrated solutions are considered feasible if both GOF_NMBE and GOF_CV < 15%. Recall that ASHRAE Guideline 14-2002 (ASHRAE 2002) suggests threshold values for NMBE and CV of 5% and 15%, respectively, for a calibration using utility bills to be deemed satisfactory. This translated into GOF_Total of 11% when using our weights.

As expected, the number of LHMC trials is a major factor; increasing this number results in better calibration. For good accuracy of fits, it would seem that 2,000–10,000 calibration trials would be needed, the lower and upper values applicable to synthetic and actual buildings, respectively. As expected, the calibration fits are excellent for synthetic buildings S1 and S2, with an upper limit to the calibration accuracy for the best trial of about 2% and for the median of the top 20 calibrated solutions of around 4%–6%. For the actual building, the fits are slightly poorer, GOF_Total values about 5%–7%. Even then, these values are half of that suggested by ASHRAE Guideline 14-2002.

Results of Set 4 of Table 3 corresponding to the actual building, A1, remain poor (median GOF_Total about 10% or higher) even when the number of trials is increased. A closer look at the previously assumed ranges (not shown in the table) of some of the parameters (specifically P10, P14, P17, and P19) led us to modify them to the values listed in Table 3. A marked improvement in the GOF_Total is obvious as a result. For example, Run 1a of set 4 had 126 parameter combinations, which satisfied our closeness of fit criteria with values on the order of 8.4%–14.5%, while 246 combinations satisfy the same threshold criteria in Run 2a with GOF_Total values around 6.0%–7.5%.This leads us to conclude that in case a calibration run yields relatively poor calibrated solutions, the analyst should consider reevaluating the stipulated range of variations of some of the influential parameters.

Figure 2 assembles six frames for building S1 in Dallas involving 2000 LHMC trials, four of which provide a general indication of how all the trials fare in terms of accuracy for electricity

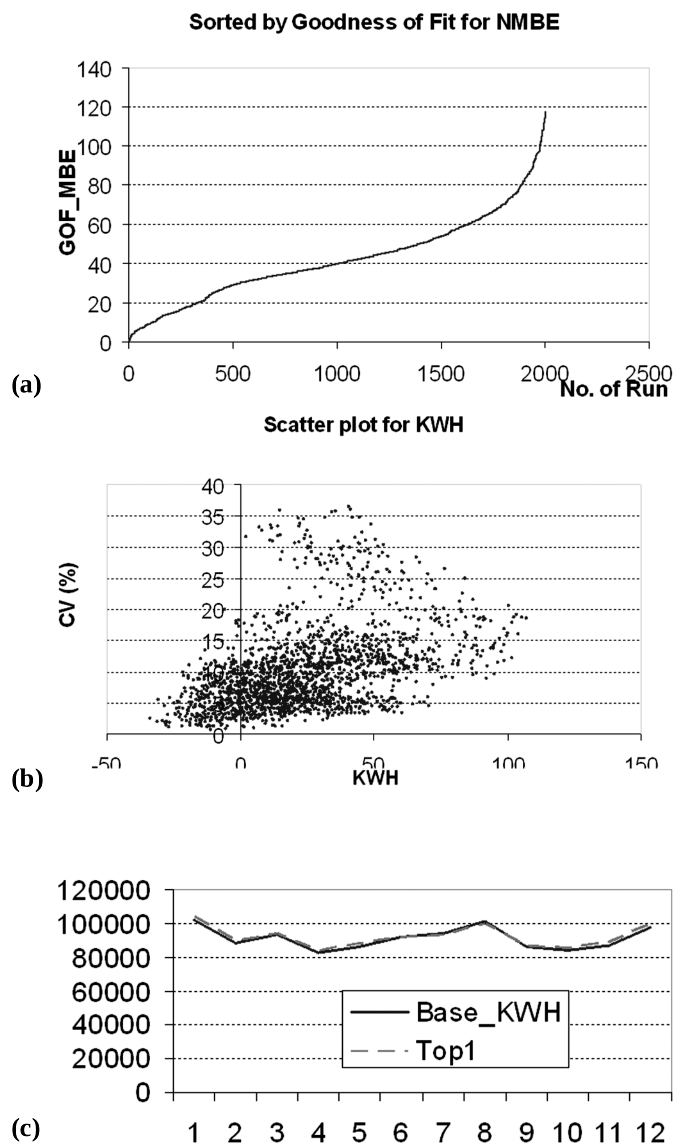


Figure 2. Results of coarse grid search for Run 3a (2000 LHMC trials) of electricity use for building S1 in Dallas, TX: (a) goodness-of-fit statistics for all trials, (b) scatter plots for all trials, and (c) monthly time series plots for the best trial.

use and demand. The last two frames, which show time series comparisons, suggest excellent fits for both kWh and kW at monthly scale for the best calibrated parameter combination. Similar time series plots for the three energy channels for the complex building S2, in Atlanta, GA, also reinforce the fact that a coarse grid search can yield very good fits to utility bill data (Figure 3).

There were a large number of strong parameters identified for A1 and S2 (about 8–12 out of 20 input parameters in this case, against only four for the synthetic building S1). Further, it was not clear as to which of the three ranges into which each parameter range was discretized

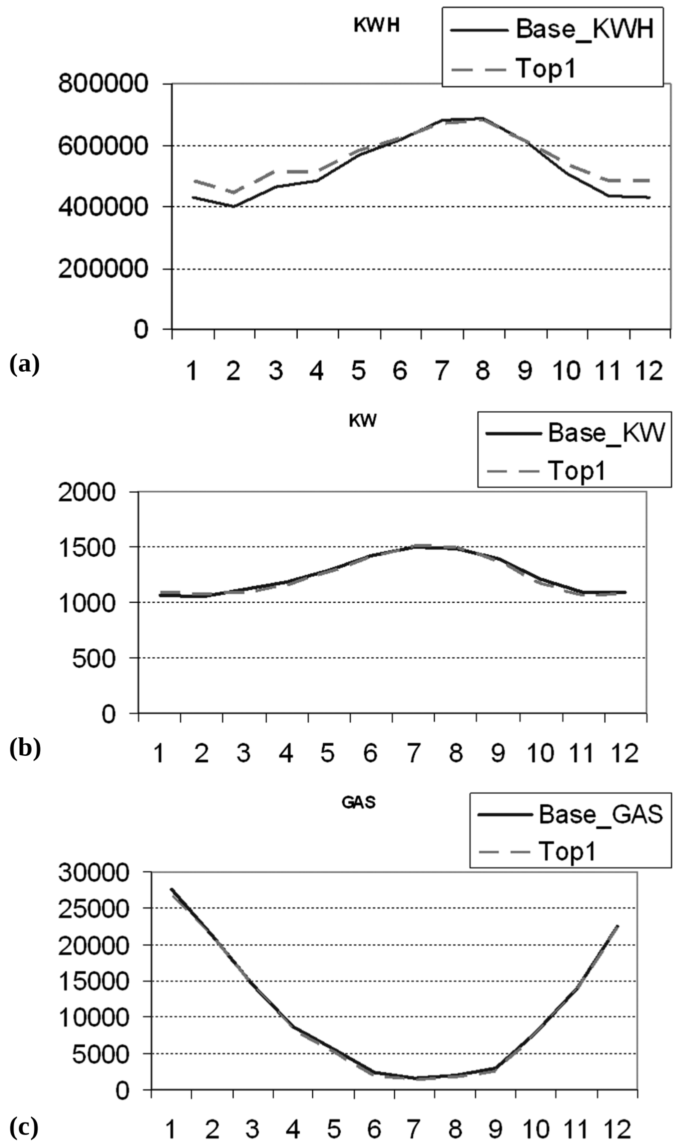


Figure 3. Results of the best trial for one stage using 3000 LHMC trials (Run 2a) for Building S2 in Atlanta, GA: (a) electricity use (kWh), (b) demand (kW), and (c) gas use (therm).

(x1-low, x2-mid, and x3-high) was preferable. Hence, we concluded that for two-stage calibration it is more practical to freeze the weak parameters rather than the strong parameters. A more basic issue is whether the two-stage calibration process is computationally more efficient than a single-stage process. Our results indicated that there does not seem to be much incentive in performing a two-stage calibration. This is illustrated by the results compiled in Table 3, where we note that though the GOF_Total values improve under the former case for the two synthetic buildings, this is not so for the actual building calibration.

Table 3. Summary of Various Sets of Runs for All Three Buildings[a]

Set	Bldg	Run #	Location	# of Trials	# of Input Parameters	# of Feasible Trials	GOF_Total of Top 20 Solutions		
							Min %	Max %	Median %
1	S1	1a	Philadelphia	100	20	21	5.33	17.03	12.91
		2a		500	20	49	2.65	9.95	8.09
		3a		2000	20	170	1.49	6.64	5.57
		2b		500/500	20/18	192	2.16	5.24	4.56
2	S1	2a	Dallas	500	20	52	3.30	12.79	9.91
		3a		2000	20	196	2.08	8.00	6.61
		2b		500/500	20/18	191	3.57	7.26	6.73
3	S1	2a	Philadelphia	500	20	24	6.36	12.10	10.07
		3a		2000	20	119	3.16	8.10	6.21
4	A1	1a	Collegeville	5000	20	126	8.40	14.47	13.16
		2a		10000	20	3873	8.67	12.12	10.12
		1b		5k/5k	20/17	156	8.83	11.50	10.45
		2b		10k/10k	20/16	678	7.73	10.96	9.81
5[b]	A1	1a	Collegeville	2500	20	128	6.54	8.80	7.89
		2a		5000	20	246	6.09	7.68	7.01
		3a		10000	20	542	6.04	7.22	6.71
		1b		2500/2500	20/13	174	6.06	9.17	7.68
		2b		5k/5k	20/15	264	6.21	8.25	7.41
6	S2	1a	Atlanta	1500	24	141	5.26	7.66	6.75
		2a		3000	24	326	3.75	6.57	6.00
		3a		5000	24	548	2.80	5.88	4.85
		2b		3000/3000	24/19	391	3.34	6.45	5.94
7	S2	1a	Dallas	1500	24	100	6.56	8.55	7.76
		2a		3000	24	210	4.71	7.58	6.74
		3a		5000	24	397	3.33	6.58	5.20
		2b		3000/3000	24/19	258	3.24	7.40	5.81

a. The range of the 24 input parameters are given in Table 2. Equal weighting factors for kWh/kW/gas utility bills are assumed, while a ratio of 3:1 has been selected for NMBE:CV values. Calibrated solutions are considered feasible if both GOF_NMBE and GOF_CV < 15%.

b. Ranges of four input variables have been readjusted.

Note: a = single-stage calibration, b = two-stage calibration (involves freezing weak parameters whose Chi-square value < 1.4).

We also looked at whether the same set of strong and weak parameters are consistently identified for each building under calibration runs with different numbers of trials and stages. We found great consistency (see Reddy et al. [2006] for details) from one run to another with different numbers of trials, leading us to conclude that the coarse grid approach is sufficiently robust in most cases to consistently identify weak and strong parameters.

On a final note, we also investigated whether our calibration process is robust enough to identify the correct strata or sub-range of all the strong input parameters. Recall that the minimum-maximum range of each parameter was discretized into three strata of equal probability assuming a triangular distribution as described in Reddy et al. (2007). We found that even when the GOF_Total is excellent (as for S1), our coarse grid search was only able to correctly identify the sub-ranges of some of the strong parameters (not *all* of them). This gets much worse for buildings A1 and S2. Hence, it would seem that being able to correctly identify the sub-ranges of all the strong parameters is highly unlikely under the previously defined scope of this study.

FINE GRID SEARCH RESULTS

The intent of the guided refined search is to start with a relatively small set of the most promising parameter vectors (say, about 20) identified from the coarse grid process and fine-tune each of these vectors one at a time so as to yield closer calibration with the monthly utility bills. This could be achieved in several ways, for example, by heuristic calibration or by treating the problem as an optimization problem, in which case such techniques as genetic algorithms or any of the traditional optimization algorithms can be used. In this study, we have chosen the mathematical optimization path, more specifically, the *fine grid approach*. This is probably the simplest, though not the most efficient computationally, and involves reducing the size of the finite increments of each of the strong variables and determining the overall minimum by performing an exhaustive search of all possible combinations. It is important to point out that this approach has been adopted in this research over the several more efficient optimization algorithms more for the sake of simplicity rather than as the one we recommend for practical use.

From Table 2, we note that of the 20 influential parameters for the simple office building category, 6 are discrete schedules while 4 are binary schedules. Handling continuous variables during an optimization process is relatively straightforward. Since a schedule is defined by a vector of 24 hourly values, one needs to adopt a procedure that allows such a vector to be manipulated so as to render it into a continuous variable (or at least into a semi-continuous variable). We propose an approximate and practical manner, which is described below.

Figure 4 is a sketch of a diurnal schedule in a typical office building with distinct occupied and unoccupied hours. For better clarity, the schedule is shown as a continuous line rather than a vector of 24 hourly values. Relatively small deviations from the base or parent schedule can be associated with the degrees of freedom, which can be of three types: **type A**, vertical scaling where the peak value can be moved either upward or downward—this would, in essence, affect the load factor; **type B**, horizontal stretching/squishing, where the number of occupied hours is increased/decreased symmetrically about noon without affecting the peak value; and **type C**, horizontal translation, where the entire profile is moved either earlier or later in the day without affecting the number of occupied hours during the day or the peak value. Following this approach, one discrete parent schedule can spawn a large number of offspring schedule combinations, any one of which could be a possible candidate during the guided refined search. It will be noted that for a binary schedule specified by either ON/OFF or 0/1, type A deviation is not applicable.

The top five calibrated solutions obtained from our coarse grid search results for Run 3a and Run 2b of the synthetic building S1, located in Philadelphia, PA, were used to initiate the fine grid search. The following four strong parameters were specifically used for fine tuning: P4—fan schedule (binary), P5—space heating temperature schedule (discrete), P17—minimum supply airflow (continuous), and P19—minimum outside air (continuous). Parameter P4 can only have type B and type C variations, while P5 can have all three. Recall that during the coarse grid generation, we had divided the range of variability of each of the 20 parameters into three sub-ranges, then identified the "most likely" sub-range for each of these parameters. The fine

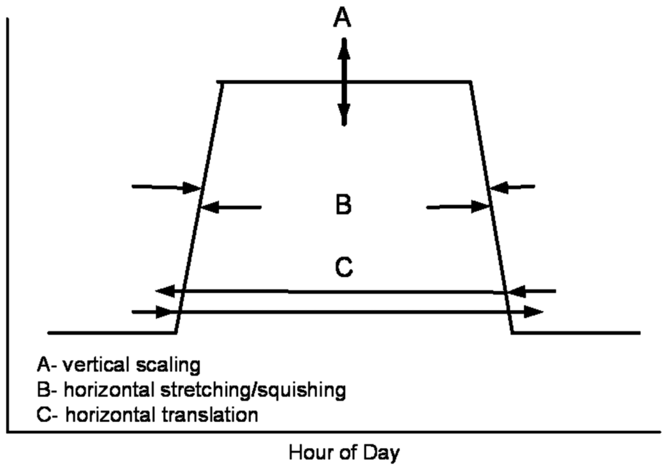


Figure 4. The three degrees of freedom or types of deviations that can be imposed on a discrete parent diurnal operating schedule to render it semi-continuous as needed during the guided refined search process.

grid starts with each of these "most likely" sub-ranges and further subdivides them into three finer ranges. Thus, P17 and P19 have been subdivided into three fine-ranges each, while P4, which is binary, will result in the generation of nine offspring schedules inclusive of the parent schedule (three for variation type B times three for variation type C). Since P5 is discrete, we have 27 offspring schedules (three values each for variation types A, B, and C). Note that we have assumed a two-hour change in variations B and C. Hence, the fine grid optimization involves running DOE-2 for (9 × 27 × 3 × 3) = 2187 times and identifying the best as the parameter vector combination that results in the least GOF_Total, i.e., the best fit to the utility bills.

DOE-2 simulations were performed 2187 times for each of the ten parameter sets identified as most plausible by the coarse grid calibration. The extent to which the GOF_Total improves by performing the fine grid calibration can be clearly noted from Figure 5. Even though the coarse grid calibration results yielded values of GOF_Total in the range of 1.5% to 4%, which were deemed to be very good, these values have been further reduced during the fine grid calibration to 1%–2%. This analysis clearly demonstrates that a fine grid calibration process can improve the fitting compared to even a very good coarse grid calibration. Whether such an improvement is warranted in the perspective of why the calibration was undertaken in the first place is addressed in the next section, which deals with uncertainty in predictions of the calibrated simulations.

UNCERTAINTY IN CALIBRATED MODEL PREDICTIONS

A background on uncertainty issues has been provided by Reddy (2006). The issue of specific relevance here relates to the accuracy with which our calibrated solutions are likely to predict energy and demand savings were specific ECMs implemented. In other words, we would like to investigate the ECM savings and its associated uncertainty as predicted by the calibrated solutions. As stated earlier, instead of relying solely on the predictions of any one calibration (even if it were to fit the utility bill data most accurately), our methodology suggests using a small set of plausible solutions. We have, somewhat arbitrarily, selected the top 20 calibrated solutions for each case, deeming this number to provide sufficient robustness while not being so large as to be unmanageable (subsequent analysis, described below, revealed that this was a sound

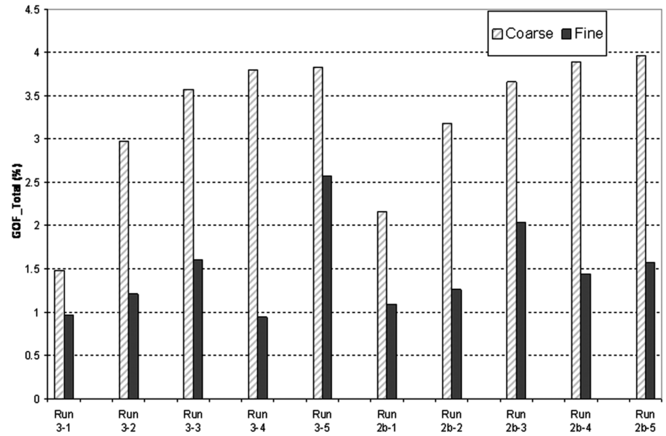


Figure 5. Improvement in calibration accuracy for ten different trials (five from Run 3a and five from Run 2b) between coarse grid and fine grid calibrations for S1, located in Philadelphia.

Table 4. Specific Parameters That Were Modified, Along with Their Original (or Baseline) Values and the New Values to Simulate the Four Sets of ECMs Considered

	Parameters	Baseline Value	Modified Value
ECM_A	P17/P19	0.65/0.30	0.30/0.10
ECM_B	P17/P19	0.65/0.30	0.20/0.25
ECM_C	P11/P12	1.67/1.0	3.0/3.0
ECM_D	P8/P9/P14	0.64/0.64/0.19	0.40/0.40/0.50

choice). The median and the interquartile standard deviation (i.e., the ten trials whose predicted savings are between the 25 and 75 percentiles) are then calculated from these 20 predictions. The effect of large deviations from any individual outlier predictions can be greatly minimized (if not eliminated) by positing that the actual value is likely to be bracketed by the interquartile standard deviation around the median value. This would provide greater robustness by discounting the effect of outlier predictions.

The predictive accuracy of the calibrated simulations has been investigated using four different sets of ECMs, some of which would result in increased energy use and demand, such as ECM_C (Table 4). This is not a major shortcoming since we are primarily interested in investigating prediction accuracy of simulated calibrations when subject to input parameter changes and whether this results in either a reduction or an increase in energy use is of secondary consequence. For example, retrofits for ECM_A involve the minimum supply airflow (P17) and the minimum outdoor air fraction (P19), both of which are strong parameters and are two of the four parameters calibrated during the fine grid process. The reference case corresponds to a value of P17 of 0.30 (reduced from 0.65) and a value of P19 of 0.1 (down from 0.3). The results are computed as the predicted percent savings in energy use and demand compared to the original building. For example, the percent savings in kWh have been calculated as

$$\%\ \text{kWh} = 100 \times \frac{\text{baseline annual kWh} - \text{predicted annual kWh}}{\text{baseline annual kWh}} \,. \tag{1}$$

A positive savings value would imply that the ECM resulted in a decrease in either energy use or demand, and vice versa. Obviously, for the two synthetic buildings, S1 and S2, the "correct" or exact values of savings are known. The conclusions reached from the numerous evaluations performed with the three buildings and their several runs were generally consistent (see Reddy et al. [2006] for complete details) and are described below.

A more accurate calibrated solution, i.e., one with lower GOF_Total, does not necessarily predict ECM savings more accurately, though there does seem to be a weak correlation (see Figure 6). Savings predictions from certain trials are clearly outliers, and the ability to detect these is clearly of great practical benefit. The proposed method of taking the interquartile range

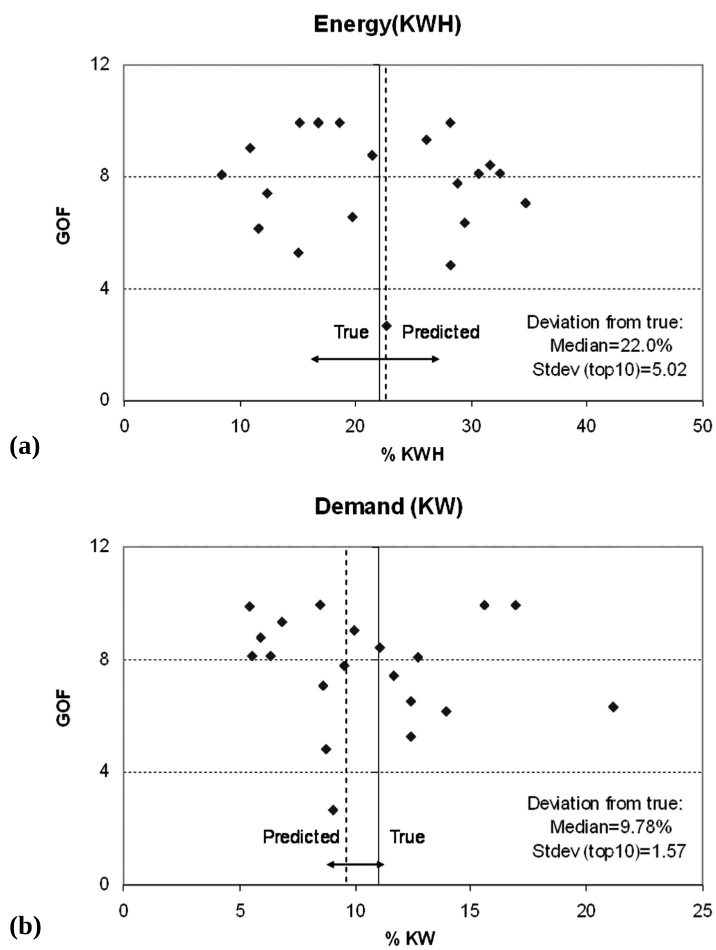
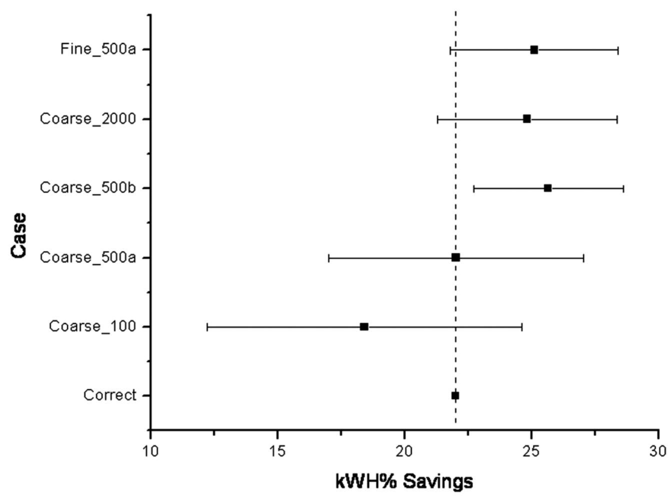


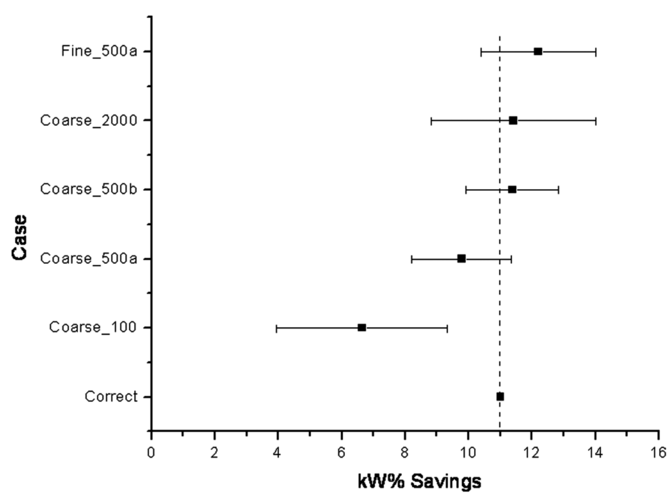
Figure 6. Predicted ECM_A savings percentage from *coarse grid* calibrated models versus their calibration accuracy expressed as GOF_Total for Run 2a (500 trials). The results correspond to synthetic building S1, located in Philadelphia. Twenty of the top calibrated solutions from Run 2a of the coarse grid search (see Table 3) have been used. The "true" or correct savings are 22% and 11% reductions in energy and demand, respectively. The median (dotted line) and the standard deviation (two-sided arrows) of the best ten interquartile predictions are also shown: (a) electricity use (kWh) and (b) demand (kW).

to calculate the uncertainty automatically results in such anomalous outliers being rejected and provides the needed robustness to our approach. Further, we find that LHMC runs with more trials tend to predict the average more accurately (as illustrated in Figure 7 for S1 and in Figure 8 and Table 5 for S2), where accuracy is determined by how close the median of the predicted values are to the correct values and whether the uncertainty bands are narrow enough to be of practical use while bounding the correct values. There are some exceptions that cannot be satisfactorily explained; for example, in Figure 8 we note a strong bias in the predicted gas use and demand savings of the calibrated solutions.

Overall, the relative uncertainty or fractional difference between "actual" and predicted have been found to be in the range of 25%–50%. Nevertheless, what is encouraging is that, in most



(a)



(b)

Figure 7. Median and interquartile standard deviation for the top 20 calibrated trials predicting savings in kWh percent and kW percent due to ECM_A for building S1. The "true" values are indicated by vertical dashed lines: (a) electricity use (kWh) and (b) demand (kW).

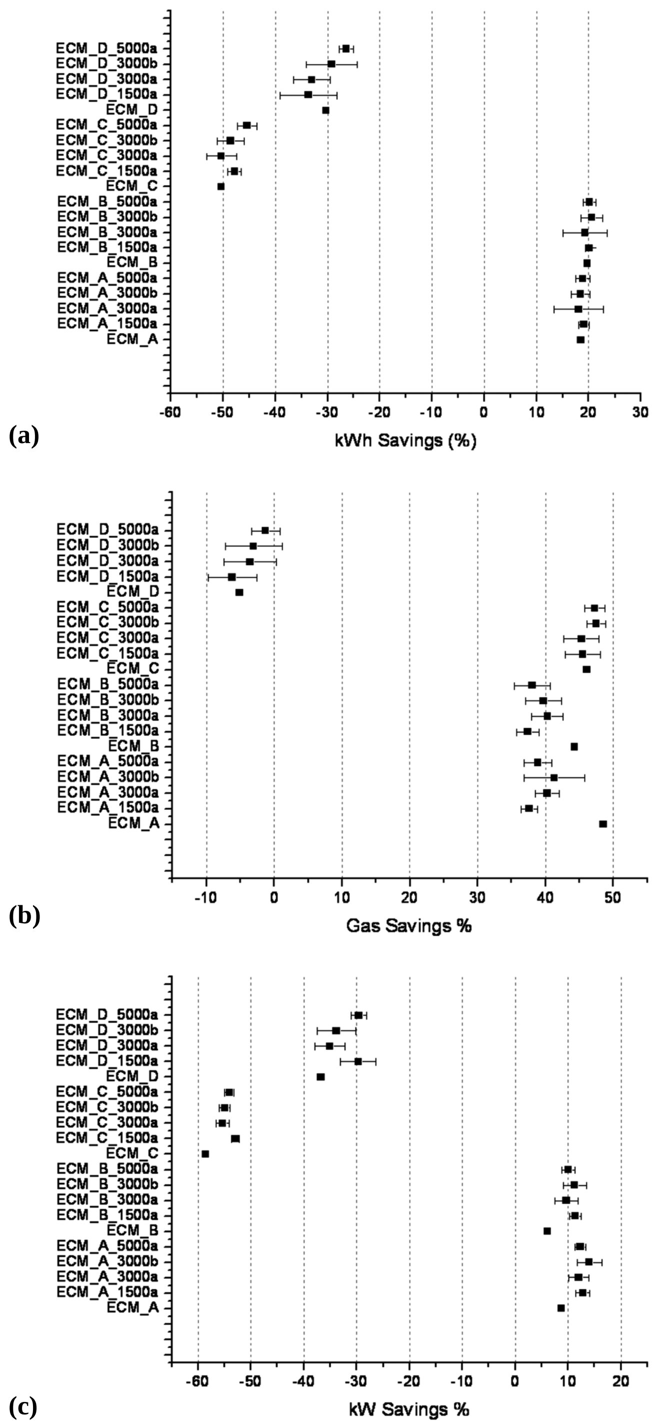


Figure 8. Energy, demand, and gas savings (in percent compared to the baseline building values) for the four ECMs predicted by the top 20 calibrated solutions from different numbers of LHMC trials. The data correspond to synthetic building S2, located in Atlanta: (a) electricity use savings, (b) gas use savings, and (c) electricity demand savings.

Table 5. ECM Savings Predictions for Synthetic Building S2 in Atlanta by Various Calibration Runs[a]

Run	# of Trials	Median GOF_Total	Predicted Savings Using Top 20 Solutions					
			kWh%		kWh%		Gas Use %	
			Median %	Standard Deviation[b]	Median %	Standard Deviation†	Median %	Standard Deviation†
Exact	—	—	**18.47**	**0.00**	**8.76**	**0.00**	**48.53**	**0.00**
1a[c]	1500	6.75	19.08	0.98	12.79	1.32	37.61	1.18
2a	3000	6.00	18.06	4.73	12.00	1.85	40.23	1.77
2b	3000	5.94	18.43	1.75	14.00	2.33	41.30	4.52
3a	5000	4.85	18.85	1.33	12.31	1.07	38.88	2.04
Exact	—	—	**19.69**	**0.00**	**6.06**	**0.00**	**44.31**	**0.00**
1a	1500	6.75	20.05	0.68	11.35	1.04	37.38	1.62
2a	3000	6.00	19.32	4.22	9.68	2.22	40.30	2.32
2b	3000	5.94	20.58	2.00	11.26	2.24	39.71	2.60
3a	5000	4.85	20.12	1.24	9.98	1.24	38.02	2.66
Exact	—	—	**–50.31**	**0.00**	**–58.63**	**0.00**	**46.10**	**0.00**
1a	1500	6.75	–47.79	1.26	–52.92	0.75	45.51	2.59
2a	3000	6.00	–50.32	2.90	–55.39	1.25	45.32	2.59
2b	3000	5.94	–48.56	2.57	–54.99	1.06	47.48	1.41
3a	5000	4.85	–45.43	1.88	–54.10	0.88	47.23	1.49
Exact	—	—	**–30.32**	**0.00**	**–36.76**	**0.00**	**–5.11**	**0.00**
1a	1500	6.75	–33.65	5.47	–29.74	3.28	–6.20	3.59
2a	3000	6.00	–32.97	3.52	–35.10	2.86	–3.57	3.84
2b	3000	5.94	–29.20	4.86	–33.82	3.64	–3.06	4.18
3a	5000	4.85	–26.42	1.33	–29.61	1.46	–1.28	2.10

a.. The “correct” savings and the savings predicted by the interquartile values of the top 20 calibrated solutions are shown.
b. Standard deviation of only those in the interquartile range (i.e., those between 25% and 75%).
c. Runs a and b refer to one-stage and two-stage calibration, respectively.

cases (but not all), the actual savings are usually contained in the range predicted, i.e., $(\text{median of predicted} \pm 2 \times \text{standard deviation})$, which would correspond to the 95% confidence level (if a normal distribution is assumed). One should expect, even in the best of cases (as for the simple synthetic building S1), that bias in predictions of calibrated simulations will be less than about 10% of the correct ECM savings predicted value and the associated standard deviation or random uncertainty to be about 3%–5% at best. More realistically, one should not rely on whole building calibrated simulation using utility bills to predict the effect of ECMs that result in less than 5% savings, while one should expect relative prediction uncertainties around 20%–30% when savings are in the order of 10%–30%. It would be advisable to be duly cautious of calibrated simulation results that predict savings from ECM less than about 10% savings (since the associated uncertainty could be 50% or more of this value).

Additionally, we found that generally (but not in each and every case), increasing the number of trials tends to result in more accurate predictions. The reduction of bias as the number of trials is increased (referred to as *consistency*) is a desirable feature in statistical analysis. Again, there are exceptions, as can be noted from Figures 7 and 8.

We have attempted to reduce the scatter by plotting the percentage savings not against a combined kWh-kW statistic (namely, the GOF_Total) but against the GOF_kWh and GOF_kW (for example, GOF_kWh includes only the NMBE and CV of the calibrated simulation trial with respect to kWh). There is no noticeable improvement in the correlation between percent savings and GOF, so the initial choice of using a combined GOF statistic (namely GOF_Total) as the basis to evaluate prediction accuracy seems appropriate.

Whether two-stage coarse grid calibration improves the savings prediction as compared to the simpler one-stage calibration has also been investigated for all three buildings (see Reddy et al. [2006] for full details). Though the accuracy improved somewhat in certain cases (see Table 5 and Figures 7 and 8 for S1 and S2), there seems to be no strong incentive for using two-stage calibrations. Instead, our conclusion is that it would be preferable to use a large number of one-stage trials.

An investigation based on building S1 was done to determine whether performing a fine grid calibration improves the prediction accuracy (see Reddy et al. [2006]). Though the prediction accuracy using the fine grid is, in almost all cases, better than the coarse grid, it is, however, not substantially better across all ECMs. While it is substantial for energy use prediction under ECM_C and ECM_D (for the latter, the “correct” savings are –5.0%, and coarse grid and fine grid simulations predict –10.1% and –4.69%, respectively), the improvements are small for ECM_A. Whether the improved accuracy is commensurate with the extra time and effort required to perform a fine grid search is uncertain and should be investigated in future studies.

We have also investigated the choice of selecting the top 20 calibrated solutions as the plausible solution set with which to make ECM savings predictions. Specifically, we have selected a sample of only the top 10 predictions instead of the top 20 as assumed above and calculated the median savings and the standard deviation of those in the interquartile range (i.e., the highest two and the lowest two predictions were discarded, and the remaining six predictions were used to compute the standard deviation). The predicted savings results using the top 10 trials are slightly more accurate in general for synthetic building S1, while they were either as good or poorer for A1 and S2. The results of the evaluation are not conclusive, and this aspect needs to be reappraised in the future using actual building data.

Finally, the effect of noise and errors in the utility bills (as would occur in practice) on the calibration process has also been evaluated for building A1. We intentionally corrupted some of the kWh and kW bills by abnormally large amounts of about 15%–30% in such a manner as not to be visually obvious (see Table 6) and performed seven different tests with four cases with only one monthly utility bill corrupted, two cases with two, and one case with three. The results are

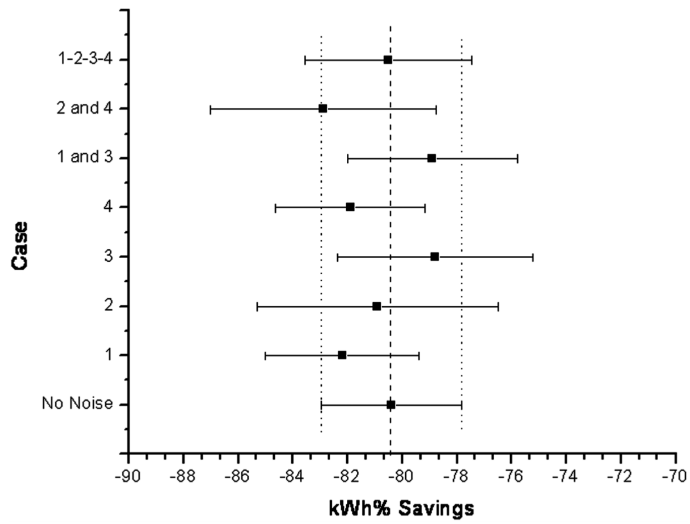
Table 6. Summary of Various Sets of Coarse Grid Runs for Actual Building A1 to Study the Effect of “Noise” in One or More of the Utility Bills

Run #	# of Trials	Affected Channel	Month	Actual Utility Bill	Modified to	Outdoor DBT (°F)	Change Point Model Parameter Affected
1	5000	kWh	Feb.	7225 kWh	8200	26	Slope
2	5000	kWh	May	5712 kWh	4500	62	Baseline
3	5000	kW	Jan.	655 kW	750	36	Slope
4	5000	kW	Apr.	788kW	700	46	Slope

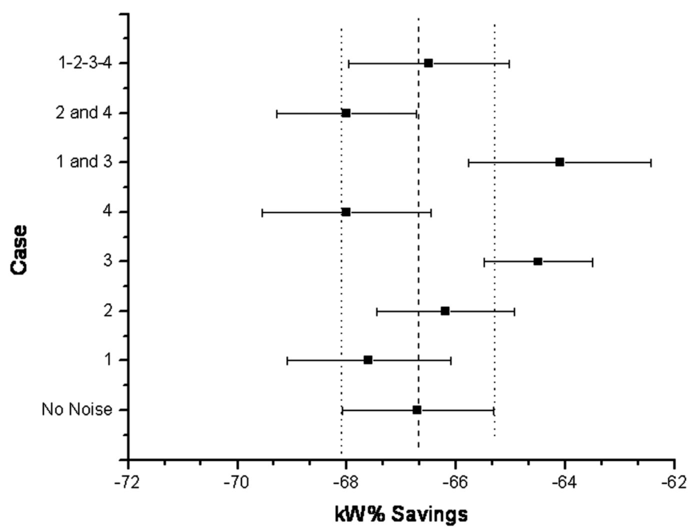
summarized in Figure 9. We find that our calibration methodology (involving prediction accuracy to various ECM implementation and associated uncertainty) was robust enough to overcome such effects since the no-noise value is bracketed by the median prediction and its uncertainty bands.

GENERAL GUIDELINES

The proposed calibration methodology has been illustrated with three case study office buildings that include two synthetic ones (an all-electric "simple" building, S1, and a dual-energy



(a)



(b)

Figure 9. Energy, demand, and gas savings (in percentage compared to baseline building values) for actual building A1 under ECM_C with different instances of utility bill corruption (see Table 6 for details). The dotted vertical line is the reference case with no corruption: (a) electricity use savings and (b) electricity demand savings.

source “complex” building, S2) and one actual building (A1), which have provided us with preliminary quantitative guidelines regarding several practical implementation issues, which are summarized below.

1. *How many influential input parameters should be selected heuristically for calibration?*

   For our “simple” buildings S1 and A1, we assumed 20 parameters, with eleven being continuous, seven being discrete schedules, and two being binary (ON/OFF). For the “complex” building, S2, the number was increased to 24 parameters. These numbers are probably too large when only utility billing data are used for calibration. The analytical calibration method suggested by Sun and Reddy (2006) links the degree of over-parametrization to the rank of the parameter matrix, and their case study indicates numbers such as five to six parameters.

2. *How many LHMC trials should be selected during coarse grid calibration?*

   Our case studies revealed that around the order of 5,000–10,000 trials should be adequate for actual buildings (though as expected the number was substantially less for the synthetic buildings studied). The number of LHMC trials should be dictated by the GOF_Total of the calibrated solutions. We suggest that the analyst use a value of 15% for GOF_NMBE and GOF_CV as threshold values for calibration in order to identify feasible solutions, though only the top 20 from this set are subsequently used. This recommendation of selecting the top 20 solutions is rather ad hoc and is meant to reduce the adverse impact of identifying a solution that, though accurate at the utility bill level, is very poor at predicting the effect of ECMs on specific end uses. Further, we found that for synthetic buildings, the better the calibration fit, the more accurate the prediction accuracy is likely to be. However, for the actual building case study, there was no clear correlation between the GOF and the prediction accuracy. We suggest from a practical point of view that the top 20 solutions should have GOF_Total < about 6% (which is twice as stringent as the value proposed by ASHRAE Guideline 14-2002). In case the first batch of LHMC runs do not meet these specifications, the analysts should either reassess the important input parameters selected and their range of variation (this was demonstrated for A1) or increase the number of trials.

3. *What value of GOF_Total of a LHMC trial can be accepted as satisfactory?*

   As stated above, one should strive for the top 20 calibrated solutions to have GOF_Total < about 6%. In case the first batch of LHMC runs do not meet these specifications, the analysts should either reassess the important input parameters selected and their range of variation (this was demonstrated for A1) or increase the number of trials.

4. *What is the advantage of performing the RSA?*

   RSA allows the analyst to identify the weak, intermediate, and strong parameters among the 20–24 influential input parameters selected and determine which one of the three discrete ranges into which the probability distribution has been discretized is most likely to bound the “correct” value of the weak and strong parameters. This is done in a statistical framework from the Chi-square statistic for each input parameter for the calibrated trials that met the GOF_Total < 15% criteria. The three case studies were fairly consistent in how they identified these from one run to another. It is obvious that this issue is important to the analyst in providing some insight about which parameters greatly affect energy use—but also from a practical viewpoint. For example, would this suggest which ECM to implement? Moreover, RSA has the ability to reduce the order of the system, i.e., decrease the number of input parameters to be tuned, and also help during the guide search grid calibration. Both these aspects have important implications from a computational viewpoint in that the number of LHMC runs can be reduced during the two-stage calibration, as discussed below.

5. *Is there any benefit in using a two-stage coarse grid calibration?*

Computational efficiency is a major issue when performing thousands of calibration trials using a detailed program such as DOE-2. Hence, it is advantageous to investigate methods to reduce the number of trials. Instead of performing a single run of, say, 10,000 trials with 20 input parameters to tune, can we obtain the same insights/results by using a two-stage calibration involving running, say, 3,000 trials first, freezing some of the weak (or strong) parameters from the RSA, then performing a second run with 3,000 trials? First, we found from our case studies that it is less ambiguous to identify the weak parameters than the strong ones and their most probable sub-range of variation. The GOF_Total values do improve, but only a little, while their prediction accuracy is not much improved. Hence, at this time, we do not recommend that two-stage calibration be done if the intention is only to reduce the computation effort.

6. *Is there a need for guided refined search calibration?*

The intent of the guided refined search is to start with a relatively small set of feasible calibrated solutions from the coarse grid calibration along with a narrower range of variability identified from RSA and to fine-tune each of these calibrated solutions one at a time so as to yield closer calibration with the monthly utility bills. As stated in the main portion of the report, this could be done in one of several ways (heuristically, by using genetic algorithms, etc.) depending on the personal preference of the analyst. The benefit that a guided search can provide in addition to the coarse grid has been investigated with building S1. The GOF_Total improves as does the ECM prediction accuracy using calibrated solutions. However, whether the improved accuracy is commensurate with the extra time and effort required to perform a guided search is uncertain and could be left to the analyst to decide based on the particular circumstance.

7. *How many top calibrated solutions should be retained for predicting ECM effect?*

Our method involves using a certain number of top calibrated solutions, rather than a single "best" calibration, to predict ECM effect and associated prediction uncertainty. The issues of how many top trials to use and how to quantify the uncertainty have been investigated. We compared the results of using the top 20 calibrated solutions (i.e., those trials with lowest GOF_Total) to those using 10 trials. Also, because of the variability of some of these estimates, we had proposed to use only the predictions within the interquartile range (i.e., within the 25–50 percentile range) and discussed how this would increase robustness since the effect of outliers can be minimized if not eliminated. The ECM savings prediction is then quantified by the median and the standard deviation of these interquartile trials. We concluded that it is statistically preferable to select the top 20 calibrated solutions for predicting ECM savings than the top 10 solutions only. However, our suggestion of using the interquartile range of the top 20 solutions in order to compute standard deviation of the predicted savings was found satisfactory for two buildings (S1 and A1) but somewhat unsatisfactory for building S2. Since the predicted savings were found to be unrealistically small (especially when the number of trials/run was increased), this aspect needs to be investigated further and refined.

8. *What is the prediction accuracy one can expect from calibrated simulations?*

Our case studies indicate that calibrated simulations can provide acceptable accuracy in predicting ECM effect if the savings are about 10% or more (with savings < 5% being extremely unreliable). The uncertainty (plus or minus two standard deviations) can be about 50% in this case, which gradually reduces to about 20% when the savings reach 30% or more.

9. *What is the effect of noise in utility bills?*

The effect of noise and errors in the utility bills (as would occur in practice) on the calibration process was also evaluated. We intentionally corrupted some of the kWh and kW bills by abnormally large amounts of about 15%–30% in such a manner as not to be obvious when visually looking at graphical plots, then we performed several different tests. We found that our calibration methodology (involving prediction accuracy to various ECM implementation and associated uncertainty) was robust enough to overcome such effects.

EXTENSIONS

This research is meant to benefit software developers in that it would specify additional capabilities to existing building energy simulation programs, which would allow calibration to be performed by practitioners with relative ease and with higher consistency. The conclusions and suggestions made above are based on the results of three case studies and should be viewed with caution since the evaluation was by no means as exhaustive as is necessary before commercial implementation. The validation of the proposed calibration approach was more of a proof-of-concept nature, and some future developments are suggested:

a. Perform exhaustive evaluation of the methodology with a large number of buildings, sizes, climates, types (hospital, hotel, etc.) so as to get a better sense of the various issues and conclusions described above. Since the matrix can become enormous, defining an optimum set of cases is not a trivial matter.
b. Initiate a round-robin test comparing this calibration method with the ones currently used, so as to clearly demonstrate the superiority of this approach and convince practitioners of its benefit.
c. Develop wizards that will prepare a short list of important input parameters to tune and identify ranges of variability and perhaps the best-guess values. This should perhaps be done for different building types (office, hospital, hotel, etc.) with perhaps finer resolution in terms of size, location, HVAC system type, etc.
d. Suggest spot and/or short-term measurements in addition to utility bills and walk-through audits, which can enhance the credibility of the calibration process and reduce the associated uncertainty of the intended result (for example, energy savings of specific end uses).
e. There is, finally, a need to better understand when the guided refined search is of added benefit or is essential. Additional extensions of this research are also listed in the final report by Reddy et al. (2006).

ACKNOWLEDGMENTS

This research was sponsored by ASHRAE TC 4.7, Energy Estimation Methods. We would like to acknowledge the Project Monitoring Subcommittee members of this project, namely, Robert Sonderegger (Chair), Bass Abushakra, Marlin Addison, Chip Barnaby, Ron Judkoff, and Vern Smith, for their timely, insightful, and constructive criticism and support during the course of this research. Several other colleagues provided advice, support, and help at different times related to different aspects of this research: David Binder, Drury Crawley, William Carroll, Charles Haas, Jeff Haberl, Srinivas Katipamula, Rob Ramirez, Jian Sun, Peter Vance, James Waltz, and Jack Wolpert.

REFERENCES

ASHRAE. 21002. *ASHRAE Guideline 14-2002, Measurement of Energy and Demand Savings*. Atlanta: American Society of Heating, Refrigerating and Air-Conditioning Engineers,Inc.

Reddy, T.A. 2006. Literature review on calibration of building energy simulation programs: Uses, problems, procedures, uncertainty, and tools. *ASHRAE Transactions* 112(1):177–96.

Reddy, T.A., I. Maor, C. Ponjapornpon, and J. Sun. 2006. Procedures for reconciling computer-calculated results with measured energy data, ASHRAE RP-1004, final report, January. Atlanta: American Society of Heating, Refrigerating and Air-Conditioning Engineers.

Reddy, T.A., I. Maor, and C. Ponjapornpon. 2007. calibrating detailed building energy simulation programs with measured data: Part I—general methodology. *HVAC&R Research* 13(2).

Saltelli, A., K. Chan, and E.M. Scott (eds.). 2000. *Sensitivity Analysis.* Chichester: John Wiley & Sons.

Saltelli, A. 2002. Sensitivity analysis for importance assessment. *Risk Analysis* 22(2):579–90.

Sun, J., and T.A. Reddy. 2006. Calibration of building energy simulation programs using the analytic optimization approach. *HVAC&R Research* 12(1):177–96.

Winkelmann, F.C., B.E. Birdsall, W.F. Buhl, K.L. Ellington, A.E. Erdem, J.J. Hirsch, and S. Gates. 1993. DOE-2 Supplement, Version 2.1E (November). Berkeley, CA: Lawrence Berkeley National Laboratory.